Am J Transl Res 2022;14(9):6604-6617
www.ajtr.org /ISSN:1943-8141/AJTR0144161

Original Article

Comparative study of disease progression for heart failure with different etiologies via time-ordered network analysis

Haoran Sun[1*], Xiuhong Li[1*], Hao Yuan[1], Chengyi Wang[1], Guangde Zhang[2], Hongbo Shi[1]

[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang Province, China; [2]Department of Cardiology, The Fourth Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang Province, China. [*]Equal contributors.

Received May 15, 2022; Accepted August 4, 2022; Epub September 15, 2022; Published September 30, 2022

Abstract: Objectives: Heart failure (HF), the primary end-stage manifestation of multiple cardiovascular diseases, has become a global epidemic with high morbidity and mortality. However, the mechanisms underlying the pathogenesis of HF with different etiologies have yet to be fully elucidated. Methods: In this study, we developed a novel method to determine the dysregulated lncRNA-mRNA regulation pairs (LMRPs) in the different causes that lead to HF. Time-ordered dysregulated lncRNA-mRNA regulation networks were constructed for comparing the HF progression initiated from different causes. Additionally, the random forest and support vector machine classification algorithm were applied to identify HF-related diagnostic biomarkers. Results: Biological functional analysis indicated that similar functions were detected at the late stage across different causes of HF, whereas different characteristics were revealed during disease progression. Specifically, the disturbance of myocardial energy metabolism might be a cause of dilated cardiomyopathy (DCM) and peripartum cardiomyopathy (PPCM), while immune response appeared earlier in hypertrophic cardiomyopathy (HCM). Inflammatory response during HCM and PPCM progression might be mediated by complement system, whereas ischemic cardiomyopathy (ICM) might be induced by cytokines. Finally, we identified several panels of diagnostic biomarkers for distinguishing HF patients of different etiologies from non-heart failure (NF) controls. Conclusions: This study revealed distinct functional characteristics during the progression of HF from different causes and facilitated the discovery of candidate diagnostic biomarkers for HF.

Keywords: Heart failure, RNA-seq, lncRNA, time-ordered level, biomarkers

Introduction

Heart failure (HF) is the end stage of various cardiovascular diseases [1], which has high morbidity rate and can be caused by various etiologies [2]. Therefore, HF has become a major public health issue [3]. Although patients of HF from different causes usually show similar symptoms, they have distinct clinical features. Nevertheless, the same treatment is usually provided to patients with HF clinically, overlooking the differences in diverse etiologies [4]. Therefore, it is important to investigate the molecular mechanisms of HF progression from different causes, which will be useful for precise clinical treatment.

Previous studies have explored the pathogenesis of different HF-inducing causes [5-11], especially the pathogenesis of dilated cardiomyopathy (DCM) and ischemic cardiomyopathy (ICM). These studies were mainly carried out at the transcriptome, proteome, and functional levels. At the transcriptome level, microarray-based data has been used to analyze the changes in gene expression in ICM and DCM. A study has reported the specific expression patterns of protein-coding RNA and long non-coding RNA (lncRNA) in the plasma of ICM and DCM patients [5]. Deep RNA sequencing of the left ventricular samples from ICM and DCM revealed that cis-gene regulation was the primary mechanism of human heart lncRNAs [6]. At the proteome level, Northern blot and Western blot analyses were performed on the ventricular tissues of DCM and ICM patients, and the results revealed that the cardiac expression of TIMP-1 and -3 transcripts and

Table 1. Summary of the dataset used in this study

GEO ID	Platform	Data type	Data set	No. of DCM	No. of HCM	No. of ICM	No. of PPCM	No. of NF
GSE141910	GPL16791	RNA-seq	Training set	166	28	0	6	166
GSE135055	GPL16791	RNA-seq	Training set	18	0	0	0	9
GSE116250	GPL16791	RNA-seq	Training set	37	0	13	0	14
GSE133054	GPL18573	RNA-seq	Test set	0	8	0	0	8
GSE46224	GPL11154	RNA-seq	Test set	8	0	8	0	8
GSE1145	GPL570	Mircoarray	Test set	27	5	31	4	11

proteins was significantly reduced in DCM and ICM, while TIMP-4 protein was significantly decreased only in ICM myocardium [7]. After proteomics and metabolomics analysis on the left ventricular tissues of ICM and DCM, some commonly perturbed pathways were detected in ICM and DCM, suggesting the overlapping molecular changes in HF [10]. At the functional level, a study reported that both cell-cell and cell-matrix adhesions were related to DCM, while immune and fibrotic reactions were typical characteristics of ICM [8]. Additionally, compared with hypertrophic cardiomyopathy (HCM), the profibrotic and metabolic pathways were specifically associated with DCM in the left ventricle tissues of mouse models [9]. The limitation of the above studies was that only two etiologies of HF were investigated, and, till now, the comparative studies across multiple different HF-inducing causes are scarce.

With the increasing HF data available publicly, many studies have systematically analyzed the different etiologies of HF [12, 13]. By identifying the significantly differentially expressed (SDE) genes from different microarray platforms, the common functional modules [12] and shared pathways [13] related to the various causes of HF have been revealed. However, all these studies were focused on the common characteristics of HF, and all the analyses were performed in a static state.

In this study, we extensively collected HF-related RNA-seq and microarray data for further analyses. First, by using a novel computational algorithm, dysregulated lncRNA-mRNA regulation pairs (LMRPs) were systematically identified in HF with different etiologies. Second, time-ordered dysregulated lncRNA-mRNA regulation networks were constructed, and the common and different characteristics were analyzed during HF progression. Finally, several panels of diagnostic biomarkers associated with different causes of HF were identified.

Materials and methods

Datasets of HF with different etiologies

HF-related RNA-seq data were extensively searched from the Gene Expression Omnibus (GEO) database, and three datasets of GSE141910, GSE135055 and GSE116250 belonging to the same sequencing platform were obtained and used as training sets. These datasets included 221 DCM patients, 28 HCM patients, 13 ICM patients, 6 peripartum cardiomyopathy (PPCM) patients, and a control group consisting of 189 non-heart failure (NF) donors. All the samples were derived from human left ventricular myocardial tissue (**Table 1**). Additionally, the RNA-seq data of GSE133054 and GSE46224 and the microarray data of GSE1145 were also acquired from the GEO and used as validation sets. GSE133054 included 8 HCM patients and 8 NF samples, GSE46224 involved 8 DCM patients, 8 ICM patients and 8 NF donors, while GSE1145 contained 31 ICM patients, 27 DCM patients, 5 HCM patients, 4 PPCM patients and 11 NF samples.

Collection of lncRNA and mRNA expression profiles

The FASTQ formats of the RNA-seq datasets were downloaded from the ArrayExpress database (https://www.ebi.ac.uk/arrayexpress/). After quality control using FastQC (version 0.11.9), the adapter sequences were removed using Trimmomatic (version 0.39) [14], and the reads were aligned to the human reference genome (GRCh37/hg19) using Hisat2 (version 2.2.0) [15]. The read counts were obtained by using featureCounts (version 2.0.1) [16]. The transcripts per million (TPM) was calculated as the expression levels of mRNAs and lncRNAs. Combat-seq [17] was applied to eliminate the batch effect between different datasets. MRNAs and lncRNAs with at least one read in more than 80% of samples were retained for

further analysis. After removing the genes containing multiple ensemble IDs, 15,163 mRNAs and 4,756 lncRNAs were retained.

For the microarray data, lncRNA expression profiles were acquired by applying an lncRNA classification pipeline based on probe sets in the Affymetrix Human Genome U133 Plus 2.0 Array [18, 19]. Probe sets corresponding to multiple mRNA or lncRNA symbol identifiers were removed. If multiple probes map to the same mRNA or lncRNA, the median for these probes was taken as the final expression value of the gene.

Identification of dysregulated LMRPs in HF with different etiologies

In this study, we hypothesized that LMRPs were abundant in NF samples and that their dysfunction would lead to the occurrence and progression of the disease. To test this, LMRPs in NF were determined with the absolute value of Pearson correlation coefficients (PCC) greater than 0.8 and p-value < 0.01. In total, 23,475 LMRPs consisting of 524 lncRNAs and 3,844 mRNAs were obtained. An lncRNA-mRNA network in NF was further constructed based on these LMRPs.

Based on the expression profile of lncRNAs and mRNAs, a dysregulated LMRP was identified by combing the following three factors: (1) the dysregulation of gene (node) expression; (2) the dysregulation of gene interaction (edge); (3) the influence of an LMRP dysregulation on the genes that directly interacted with it in NF lncRNA-mRNA networks. Specifically, the node score in an LMRP was calculated by equations (1) and (2) [20]:

$$S_{node} = F_Z(Z) = P\{Z \le z\} = 1 - \sqrt{\frac{2}{\pi\sigma^2}} \int_0^{\infty} e^{-(x^2/2\sigma^2 + \lambda z/x)} dx,\ z \ge 0 \quad (1)$$

$$z = (-\log_{10} p\text{-}value) \cdot |\log_2 FC| \quad (2)$$

Where $F_Z(Z)$ was the cumulative distribution function (CDF) of z-statistics, *p-value* was the p value reflecting the significance of differential gene expression calculated by DESeq2, and *FC* was the corresponding fold change of the gene expression. λ = ln10 and σ^2 were the variance of fold change (FC). Then, the edge score in an LMRP was computed by equation (3)-(5) [21]:

$$S_{edge} = 2\varphi|D| - 1 \quad (3)$$

$$D = \frac{F(r_{case}) - F(r_{control})}{\sqrt{\frac{1}{n_{case} - 3} + \frac{1}{n_{control} - 3}}} \quad (4)$$

$$F = \frac{1}{2} \ln \frac{1 + r}{1 - r} \quad (5)$$

Where φ was the CDF of the normal distribution. r_{case} and $r_{control}$ were the PCCs of gene expression in the case and control samples, respectively. n_{case} and $n_{control}$ were the sample numbers of case and control, respectively. F was the Fisher transformation function, applied to make the data approximately follow normal distribution [22].

Next, the influence score of an LMRP on the expression of interacting genes was calculated by equation (6):

$$S_{influence} = S_{influence}(\text{mRNA}|\text{lncRNA}) + S_{influence}(\text{lncRNA}|\text{mRNA}) \quad (6)$$

Here, $S_{influence}$(mRNA|lncRNA) was the influence score of the lncRNA in the LMRP on mRNAs that directly interacted with the lncRNA, and it was defined as 1-(p-value), where the p-value was calculated by Fisher's exact test, which reflected the significant level of the mRNAs directly interacted with the lncRNA that was enriched in SDE mRNAs Benjamini-Hochberg (BH) adjusted $p < 0.05$ and $|\log_2 FC| > 1$. Similarly, $S_{influence}$(lncRNA|mRNA) was computed. Finally, an LMRP dysregulated score was computed by combining its node score, edge score and influence score:

$$S_{LMRP} = \sum_{node \in LMRP} S_{node} + S_{edge} + S_{influence} \quad (7)$$

Furthermore, to evaluate the significance of each LMRP score, a permutation test was performed. We randomly selected an lncRNA and an mRNA to construct a random LMRP and repeated this process 10,000 times. All the random LMRP scores were calculated through the above equations and the null distribution of LMRP scores was acquired. For an observed LMRP, the empirical p-value was defined as the

proportion of randomly obtained scores larger than the observed LMRP score as follow:

$$p = (Number\ of\ S_{random} > S_{LMRP})/10000 \quad (8)$$

Here, the LMRPs with $p < 0.05$ were considered as dysregulated LMRPs.

Construction of time-ordered background lncRNA-mRNA regulation networks (TO-BLMRN) in NF

The LMRPs in NF formed an lncRNA-mRNA regulation network, and the largest component with 23,372 LMRPs was defined as the background lncRNA-mRNA regulation networks in NF, which included 472 lncRNAs and 3,746 mRNAs. TO-BLMRN in NF was generated based on a previous research by using breadth-first search algorithm (BFS) and initial nodes [23]. The initial nodes were selected as the nodes with the smallest dysregulated score. Since a node might be involved in multiple LMRPs, we defined the dysregulated score of a node as the mean value of the LMRPs dysregulated scores in which it involved. Importantly, a single-gene-based network was transformed to a LMRP-based network for facilitating the subsequent analysis. A LMRP was considered at a certain time-ordered level if any node in it first appeared at that level. Meanwhile, this LMRP was deleted at the later levels.

Collection of mRNAs and lncRNAs related to HF with different etiologies

Disease-related mRNAs and lncRNAs were extracted from DisGeNET (version 7.0) [24] and LncRNADisease (version 2.0) [25] databases, respectively. In addition, disease-related lncRNAs were also manually collected from PubMed. For each article, the abnormal expression of lncRNAs in different etiologies of HF was manually confirmed. Finally, we obtained 828 mRNAs and 7 lncRNAs, 357 mRNAs and 12 lncRNAs, 103 mRNAs and 9 lncRNAs associated with DCM, HCM, and ICM, respectively. No mRNAs or lncRNAs associated with PPCM were found.

Identification of potential diagnostic biomarkers

Genes highly associated with disease initiation were screened using the random forest supervised classification algorithm [26], which was performed by the R 'RandomForest' package. The importance score of each gene was calculated using the out-of-bag sample by permutation test, and two-thirds of the genes with the highest scores were retained each time until a balance was reached between the classification accuracy and the number of genes. The support vector machine (SVM) classification model was applied to identify candidate diagnostic biomarkers. Furthermore, the classification efficiency was measured via the classification accuracy and the area under the receiver operating characteristic curve (AUC) using 5-fold cross-validation. The processes were implemented by R 'e1071' and 'pROC' packages.

Moreover, in order to evaluate the classification efficiency of LMRPs as diagnostic biomarkers, the regulatory strength of a LMRP for a sample k was calculated utilizing the measure of 'covariability' by the following equation [27]:

$$C^k_{lnc,mRNA} = \frac{e^k_{lnc} - \mu_{e_{lnc}}}{\sigma_{e_{lnc}}} \cdot \frac{e^k_{mRNA} - \mu_{e_{mRNA}}}{\sigma_{e_{mRNA}}} \quad (9)$$

Where e^k_{lnc} and e^k_{mRNA} represented the expression level of the lncRNA and mRNA in sample k, respectively. $\mu_{e_{lnc}}$ and $\mu_{e_{mRNA}}$ denoted the average, while $\sigma_{e_{lnc}}$ and $\sigma_{e_{mRNA}}$ denoted the standard deviation of the lncRNA and mRNA across samples. The identification of the LMRPs diagnostic biomarkers was performed as described above.

Results

Global gene expression patterns in HF with different etiologies

RNA-Seq data of DCM, HCM, ICM and PPCM from the same platform were collected from the GEO database. The ComBat-Seq method was adopted to eliminate the batch effects, and principal component analysis (PCA) was performed. The data presented in **Figure 1A** and **1B** indicated that the batch effects were clearly removed. As shown in **Figure 1C** and **1D**, we found that the overall expression of lncRNAs was lower than the expression of mRNAs in HF with different etiologies. About 70% of lncRNAs had TPM values between 0 and 1, but only less than 1% of them had TPM values > 20, consistent with the current understanding that the

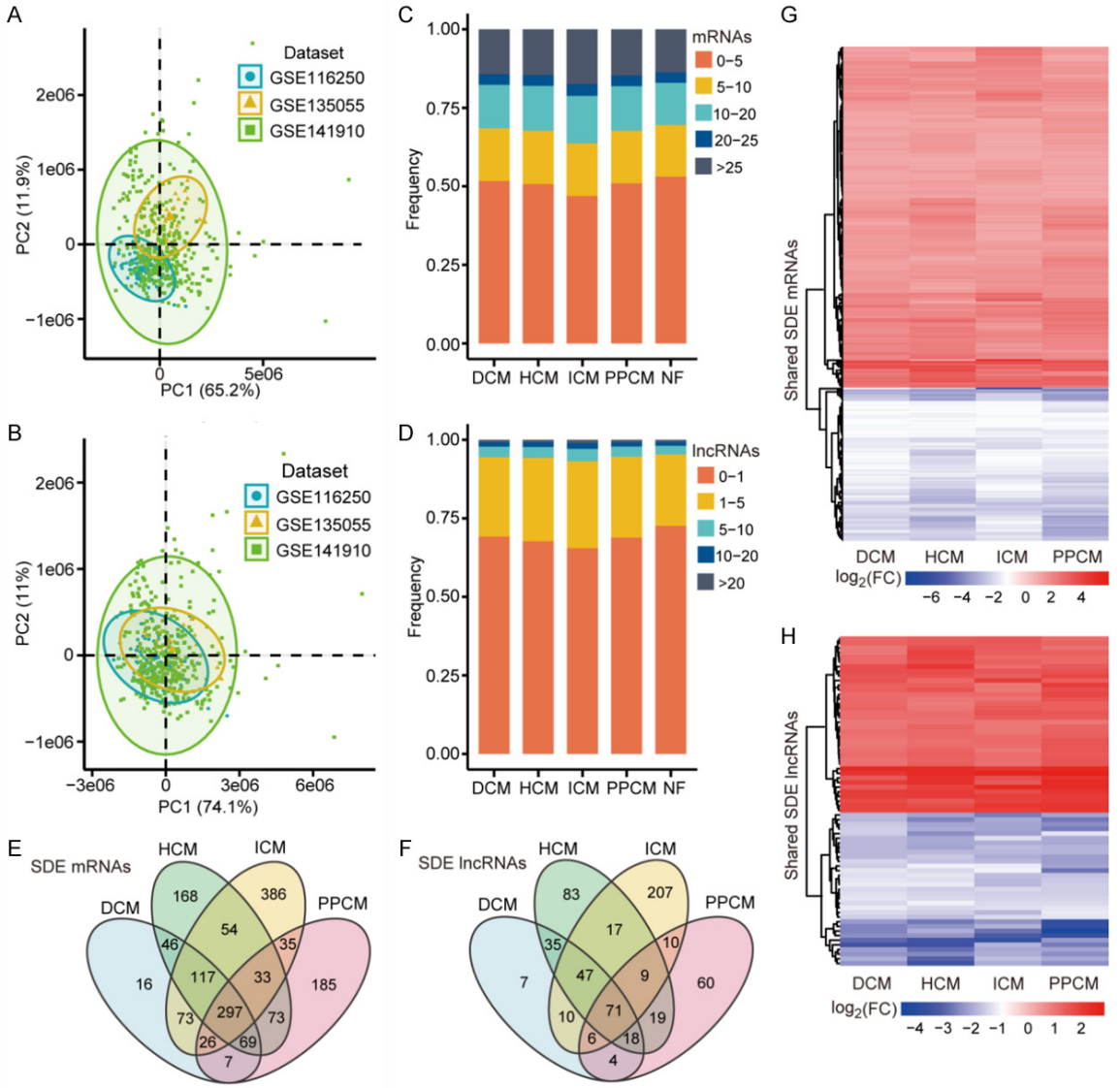


Figure 1. Global transcriptome expression patterns in heart failure (HF) with different etiologies. A, B. The scatter plot before and after removing batch effects using principal component analysis (PCA). The points in the scatter plot visualize the samples by the first two primary elements (PC1 and PC2). C, D. The distribution of TPM values for mRNAs and long noncoding RNAs (lncRNA). E, F. Venn diagrams of differentially expressed (SDE) mRNAs and SDE lncRNAs. G, H. Heatmaps of shared SDE mRNAs and SDE lncRNAs in HF with different etiologies.

expression values of protein-coding genes in different tissues were higher than those of lncRNAs [28].

Subsequently, SDE genes between HF patients with different etiologies and controls were identified at the significance level of BH-adjusted $p < 0.05$ and $|\log_2 FC| > 1$ (**Figure 1E** and **1F**). As a result, 297 SDE mRNAs and 71 SDE lncRNAs were shared. Importantly, the up- and down-regulation of all these shared SDE gene expression were consistent among the four diseases (**Figure 1G** and **1H**; Supplementary Table 1). Meanwhile, we found that 16 SDE mRNAs and 7 SDE lncRNAs were specific for DCM, while 168 SDE mRNAs and 83 SDE lncRNAs were specific for HCM. The numbers of specific SDE mRNAs and SDE lncRNAs for ICM were 386 and 207, respectively, while 185 SDE mRNAs and 60 SDE lncRNAs were specific for PPCM.

Dysregulated LMRPs in HF with different etiologies

The dysregulated LMRPs across different etiologies of HF were identified using our method.

Consequently, 2,151 dysregulated LMRPs including 943 mRNAs and 68 lncRNAs were identified in DCM. A total of 1,689 dysregulated LMRPs consisting of 428 mRNAs and 86 lncRNAs were identified in HCM. In ICM and PPCM, 2,221 dysregulated LMRPs including 586 mRNAs and 157 lncRNAs, and 1,652 dysregulated LMRPs consisted of 471 mRNAs and 81 lncRNAs were obtained, respectively (Supplementary Table 2).

More importantly, the dysregulated LMRPs were validated by several approaches and were compared with the traditional method. We first examined the proportion of SDE genes and disease-related genes (DRGene) in dysregulated LMRPs. We found that the proportion of SDE genes in the dysregulated LMRPs was significantly higher than that of NF LMRPs (hypergeometric test, $P < 2.2\times10^{-16}$ for the four diseases). Similarly, the proportion of DRGenes in dysregulated LMRPs was significantly higher than that of NF LMRPs, except for PPCM (DCM: $P = 4.37\times10^{-3}$, HCM: $P = 1.11\times10^{-2}$, ICM: $P = 8.57\times10^{-6}$). We then employed the traditional method to acquire dysregulated LMRPs. SDE lncRNA-SDE mRNA pairs with |PCC| > 0.8 and p-value < 0.01 were identified (Supplementary Table 3). However, the proportion of DRGenes in these dysregulated LMRPs was not significantly higher than that of NF (DCM: $P = 0.495$, HCM: $P = 0.104$ and ICM: $P = 0.098$). Additionally, the biological functions of the dysregulated LMRPs were investigated. The significantly enriched gene ontology biological process (GO BP) terms were acquired with $p < 0.01$ using DAVID (version 6.8) [29] (Supplementary Tables 3 and 4). As demonstrated in **Figure 2**, the traditional method only identified the processes involved in immune, inflammatory, and myocardial remodeling responses, whereas our method obtained more functional information associated with HF.

TO-BLMRN in NF

To construct TO-BLMRN in NF, DENND1B and EXOSC1 were chosen as the initial nodes because both genes were in the top 10% lowest dysregulated scores in the four diseases (for details, see Materials and methods). The TO-BLMRN was then obtained using the BFS algorithm. Furthermore, the single gene-based networks were converted into LMRP-based networks. The TO-BLMRN included 10 time-ordered levels (denoted L1 to L10). As demonstrated in **Figure 3A** and Supplementary Table 5, the dysregulated LMRPs in the diverse etiologies of HF were mapped to the TO-BLMRN, and we found that most of the dysregulated LMRPs were distributed at L3-L9 levels.

When we constructed the TO-BLMRN, we hypothesized that the initial nodes should be relatively stable; otherwise, it would cause the instability of the entire TO-BLMRN. To prove this, we calculated the ratio of dysregulated LMRPs in DCM to all LMRPs at their respective levels and repeated this process for HCM, ICM and PPCM. As shown in **Figure 3B**, as the time-ordered level increased, the ratio of dysregulated LMRPs gradually increased, especially at L8 and L9 levels. We further investigated the distribution of dysregulated scores of LMRPs at ten time-ordered levels (**Figure 3C**). Similarly, the LMRP dysregulated scores increased gradually as the level increased from L1 to L9, which was consistent with our hypothesis.

Additionally, we investigated the stability of the TO-BLMRN. Since there were only two genes at the L1 level except for DENND1B and EXOSC1, we randomly selected two genes at the L2 level of the original TO-BLMRN as initial nodes and calculated the differences in each level for new TO-BLMRN against the original one. The process was repeated 10 times, and the results were shown in **Table 2**. We found that most changes manifested at the following two time-ordered levels, probably because TO-BLMRN was constructed based on LMRPs and that the initial nodes were taken from L2 level. Moreover, the means and standard deviations (SDs) of overall level changes in new TO-BLMRNs were small.

Comparison and analysis of HF progression with different etiologies

To systematically investigate the common and different characteristics during HF progression with different etiologies, the mRNAs of dysregulated LMRPs in each level were employed for functional enrichment analysis using DAVID [29]. Significantly enriched GO BP terms were acquired with $p < 0.01$ (Supplementary Tables 6, 7, 8, 9). **Figure 4** showed the terms that were typically related to HF initiation and progression. We found that, in the advanced stage (L8-L9) of the diseases, most of the

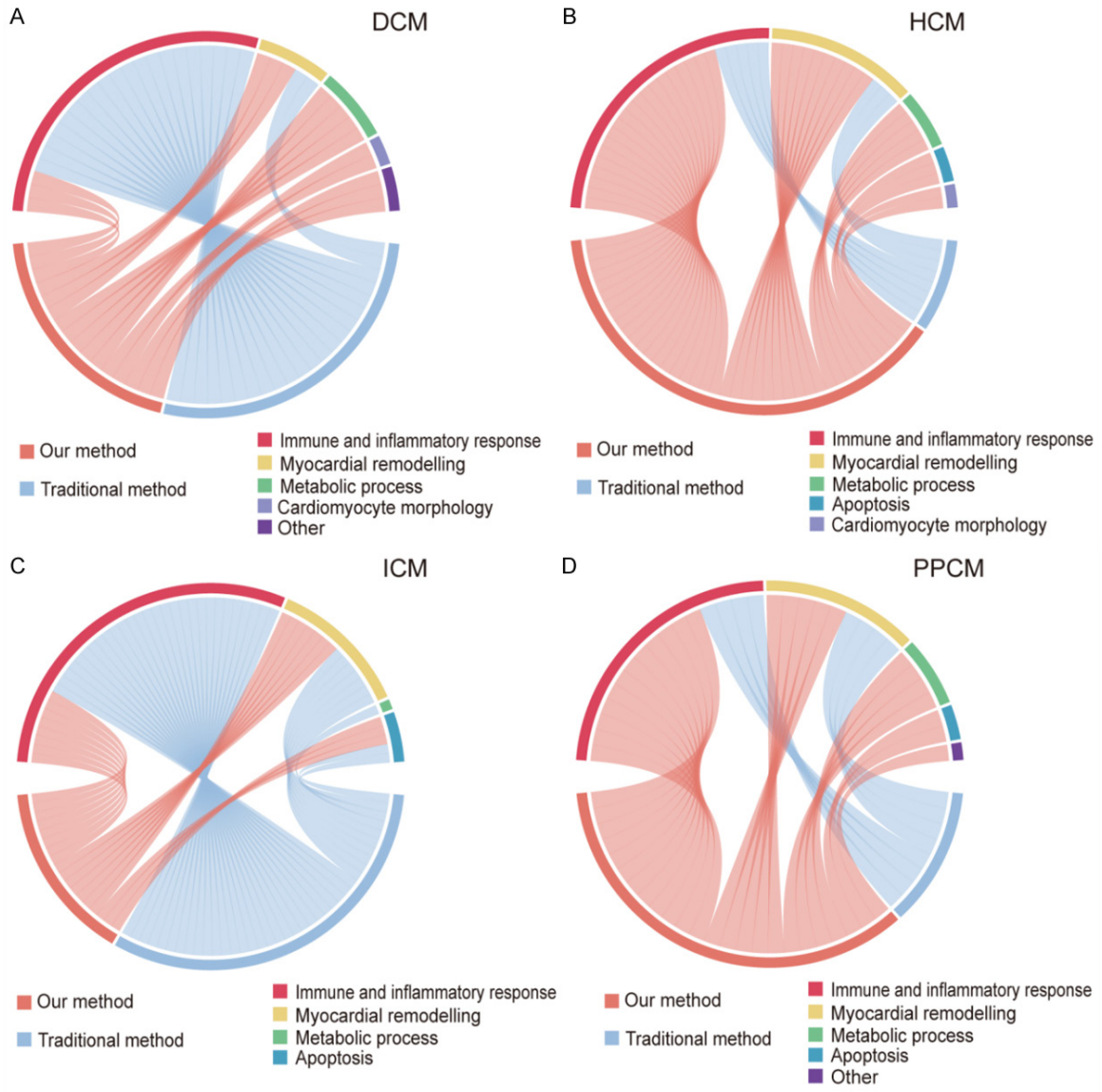
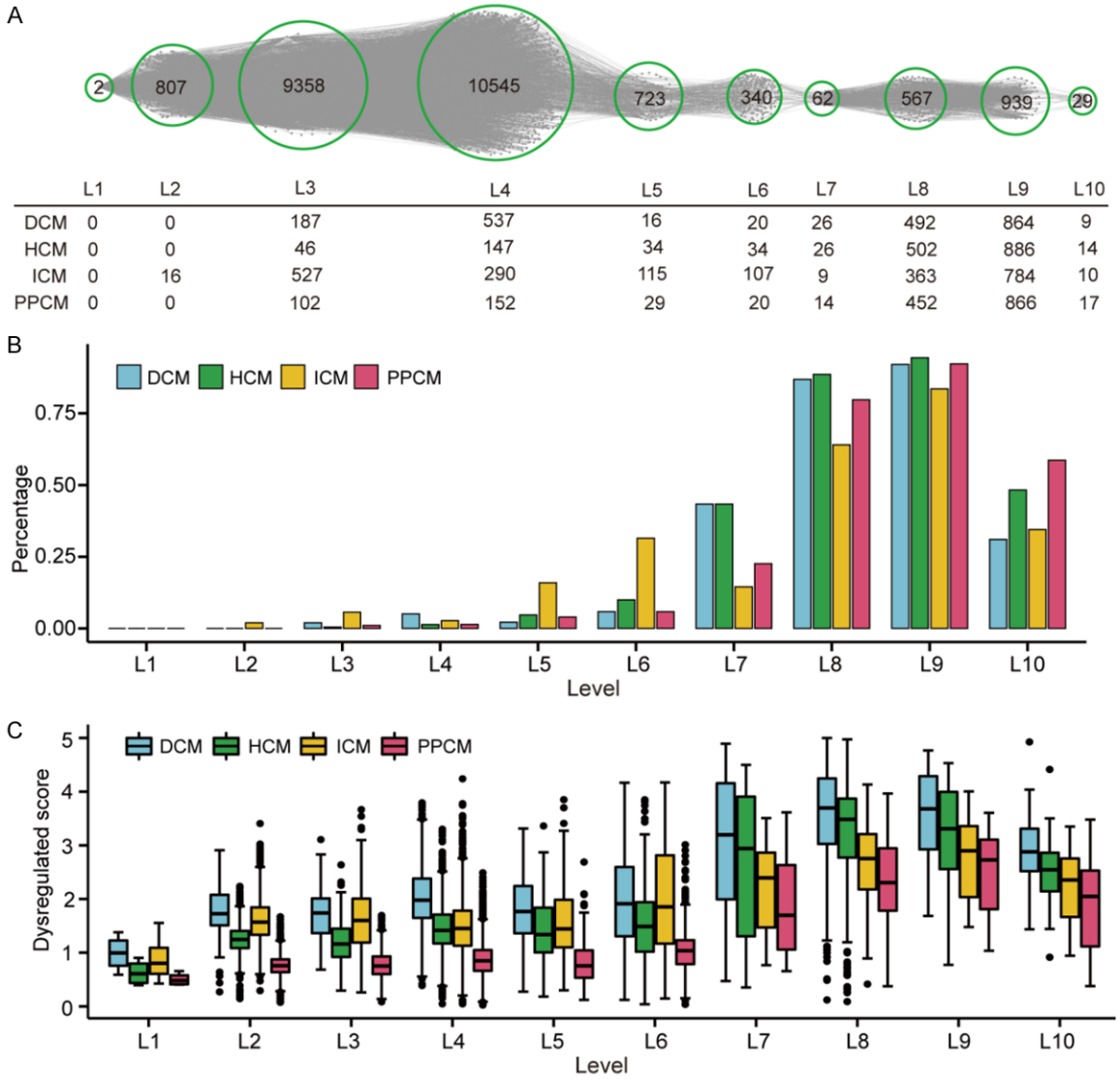


Figure 2. Comparison of our method and traditional method in biological functions. To clearly demonstrate the results, the functional categories typically associated with HF initiation and progression were selected, and similar functions were merged and classified into six categories: immune-inflammation, myocardial remodeling, metabolic progress, apoptosis, cardiomyocyte morphology and others.

immune and inflammation-related functions and the myocardial remodeling-related functions were overrepresented across the four causes of HF. For example, 'leukocyte and granulocyte migration and chemotaxis', 'cytokine production', and 'regulation of immune system process' are associated with immune and inflammation, while 'extracellular matrix organization', 'tissue remodeling', and 'angiogenesis' are related to myocardial remodeling. We also compared the pathway enrichment among the four causes of HF. Specifically, for DCM, as shown in **Figure 4A**, the 'cellular protein and macromolecule catabolic processes' and 'regulation of glucose transport' were significantly enriched at the initial stage of DCM (L3 level), and their activities continued to L4. Functional categories of 'regulation of actin polymerization or depolymerization' related to cardiomyocyte morphology were enriched in L5. At the progression stage of DCM (L6 level), the significantly enriched functions could be classified into two categories: cell apoptotic and immune response.



	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
DCM	0	0	187	537	16	20	26	492	864	9
HCM	0	0	46	147	34	34	26	502	886	14
ICM	0	16	527	290	115	107	9	363	784	10
PPCM	0	0	102	152	29	20	14	452	866	17

Figure 3. Time-ordered background lncRNA-mRNA regulation networks (TO-BLMRN) in non-heart failure (NF). A. Global overview of the TO-BLMRN and summary of dysregulated lncRNA-mRNA regulation pairs (LMRPs) in HF with different etiologies. Time-ordered levels of the TO-BLMRN are denoted as L1-L10. The number in a circle indicates the number of LMRPs. B. The proportion of dysregulated LMRPs at each level in HF with different etiologies. C. The dysregulated score distribution of LMRPs across 10 time-ordered levels.

In contrast, for HCM, the immune-related functions appeared earlier at L4 level, such as 'immune response-regulating signaling pathway', 'Fc receptor signaling pathway' and 'positive regulation of I-kappaB kinase/NF-kappaB signaling' (**Figure 4B**). Simultaneously, functions associated with cardiomyocyte morphology were overrepresented at L4, such as 'cytoskeleton organization' and 'actin filament-based process'. At L6 level, the significantly enriched functions of HCM involved cell apoptotic, immune, and inflammatory responses.

As for ICM, functional categories associated with 'response to zinc ion' were significantly enriched in L5 level (**Figure 4C**). Additionally, apoptotic signaling pathway, inflammation and immune responses were overrepresented in L6. Finally, for PPCM, the 'cellular protein and macromolecule catabolic processes' were significantly enriched in L3 (**Figure 4D**). 'Cellular response to zinc ion' was a major process at L5, and a large amount of inflammation and immune responses, including 'complement receptor-mediated signaling pathway' and

Table 2. Summary of level sequence changes with different initial nodes

Initial node	Mean of level change	SD of level change	No. of not changed	No. of cross 1 level	No. of cross 2 levels	No. of cross multiple levels
SCYL3, SMIM20	0.732	0.791	11266	7090/-2	5014/0	0/0
ANKIB1, BRK1	0.876	0.803	9158	7942/-2	6270/0	0/0
ALS2, COA8	0.9	0.814	9022	7646/-2	6702/0	0/0
RECQL, TRAPPC2B	0.752	0.786	10831	7496/-2	5043/0	0/0
KMT2E, TIMM23	0.786	0.791	10348	7683/-2	5339/0	0/0
MYCBP2, MRPS21	0.999	0.788	7261	8861/-2	7248/0	0/0
AGK, NCBP2AS2	0.983	0.798	7643	8491/-2	7245/0	0/0
PAFAH1B1, GTF2H5	0.755	0.786	10788	7522/-2	5060/0	0/0
CSDE1, PSMB3	1.033	0.815	7386	7829/-2	8155/0	0/0
REV3L, ISY1	0.984	0.809	7830	8077/-2	7463/0	0/0

Note: The positive and negative number indicate the node level changes to the previous or the following level, in the new time ordered background mRNA-long noncoding RNA (lncRNA) regulation networks (TO-BLMRN) against the original one. Zero represents no change between the two TO-BLMRNs. For example, if a node belongs to L2 in the original TO-BLMRN and is classified to L1 in the new TO-BLMRN, it is considered that the change number is -1.

'granulocyte and leukocyte migration' were overrepresented in L6.

In summary, four etiologies of HF showed the enrichments of cellular processes with different biological functions across L3-L6 levels. The finding that the 'cellular protein and macromolecule catabolic processes' were significantly enriched in the initial stage of DCM and PPCM suggested that the occurrence of DCM and PPCM might be associated with myocardial energy metabolism. However, the functional categories related to cardiomyocyte morphology were overrepresented in the early stage of DCM and HCM, suggesting that the abnormal morphology of cardiomyocytes was likely to be an early feature of the two diseases. Additionally, our results showed that the 'response to zinc ion' was possibly related to ICM and PPCM, and the inflammatory response during HCM and PPCM progression was probably mediated by the complement system, while ICM might be induced by cytokines. Furthermore, compared to the other three causes of HF, the functions associated with 'immune response' appeared earlier in HCM.

Identification of the potential diagnostic biomarkers in HF with different etiologies

In our study, two types of diagnostic biomarkers were identified based on the identified dysregulated LMRPs. One was single type molecule-based biomarkers, and the other was LMRP-based biomarkers. They were then used for differentiating HF patients from the NF controls (denoted as HF-NF). Since HF patients with different etiologies were similar in the late stage of the diseases, characterized by myocardial remodeling, the dysregulated LMRPs in L1-L7 levels were selected for diagnostic biomarkers identification.

To identify single type molecule-based diagnostic biomarkers, 786 DCM dysregulated LMRPs, including 758 mRNAs and 26 lncRNAs, were evaluated. Using our method (for details, see the Materials and methods), the optimal DCM-NF biomarkers were identified. As shown in **Table 3**, two panels of DCM-NF biomarkers were defined by three mRNAs (COPZ2, CPED1 and TGFBR2) and four lncRNAs (AC244090.1, SNHG3, LINC01091 and AC004687.1). Similarly, two panels of HCM-NF biomarkers defined by three mRNAs (PLAGL1, SCG2 and HCLS1) and three lncRNAs (LINC01128, LINC00894 and AC078850.1) were obtained. Two panels of ICM-NF biomarkers defined by two mRNAs (METRN and EIF4EBP1) and three lncRNAs (ANKRD10-IT1, AC124798.1 and AC027279.1) were obtained. Two panels of PPCM-NF biomarkers defined by two mRNAs (IFT43 and GTF3A) and three lncRNAs (PCBP1-AS1, LINC01128 and LINC01554) were obtained.

For the identification of LMRP-based diagnostic biomarkers, the 'covariability' method was applied. In the same way, four panels of diagnostic biomarkers were defined by three LMRPs (AC244090.1/COPZ2, AC244090.1/GGT5 and

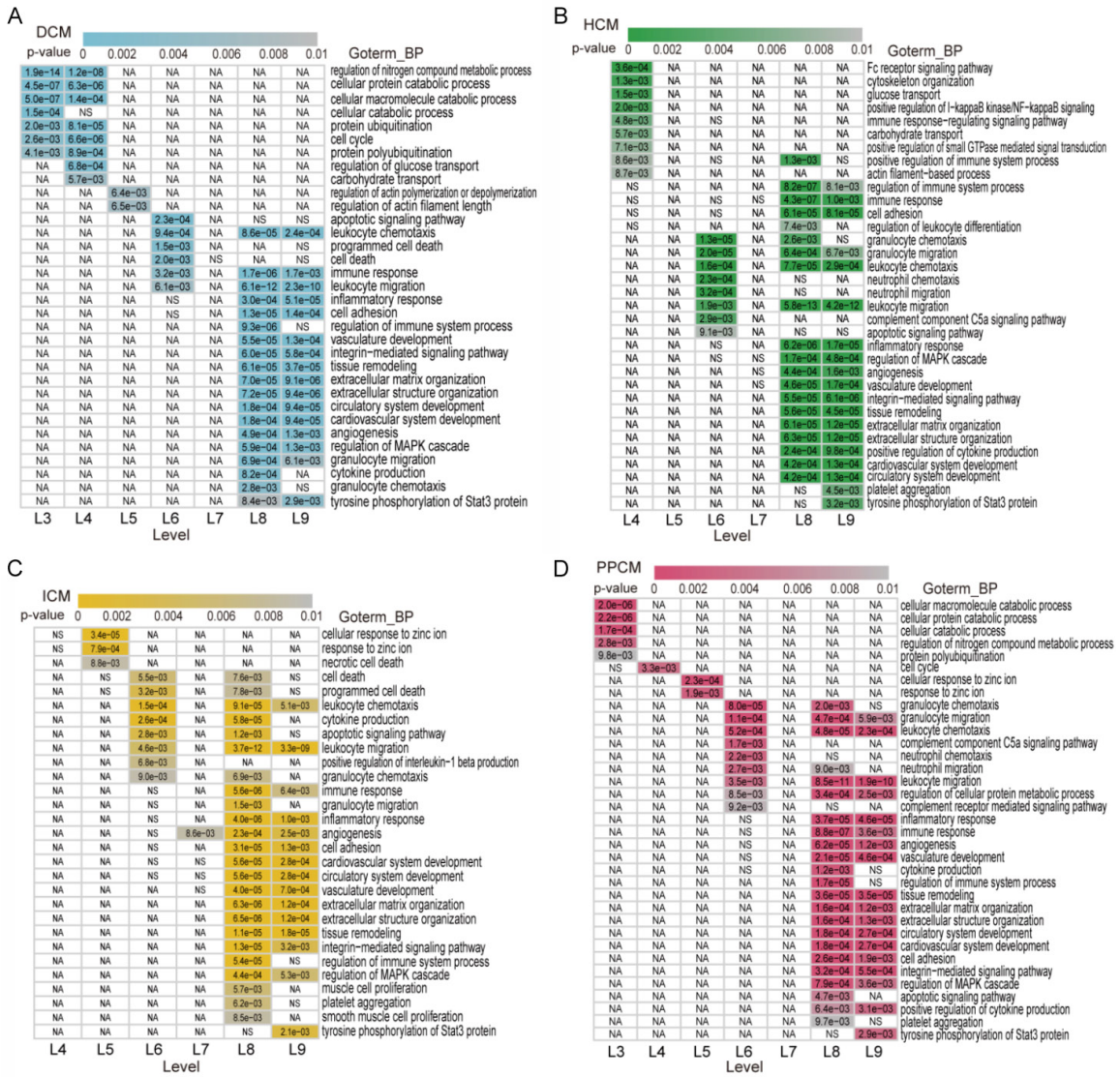


Figure 4. Dynamic changes of biological functions of HF with different etiologies. Significantly enriched gene ontology biological process (GO BP) terms typically related to HF initiation and progression were displayed. Coloring was performed based on *p*-values. 'NS' means non-significant ($P > 0.01$), while 'NA' means that the mRNAs are not enriched in this term.

FGD5-AS1/CPED1) for selected DCM-NF, while four LMRPs (SNHG6/IFT43, FGD5-AS1/TACC1, FGD5-AS1/TGFBR2 and LINC00342/PLAGL1) were defined for HCM-NF. ICM-NF was represented by two LMRPs (TP53TG1/UQCRB and RAB11B-AS1/EIF4EBP1), and PPCM-NF by three LMRPs (ZFAS1/RBM47, AL513327.1/NCBP3 and AC002350.2/DCBLD2). The accuracies and AUC values in the training and validation sets using 5-fold cross-validation were shown in **Table 3**. The results revealed that the biomarkers we identified could effectively distinguish HF patients with different etiologies from the NF controls.

Discussion

In the current study, we investigated the common and distinct functional characteristics during HF progression with different etiologies. By using a novel computational approach, the dysregulated LMRPs in the different causes of HF were identified, and furthermore, time-ordered dysregulated lncRNA-mRNA regulation networks were constructed, from which several new biomarker panels potentially useful for HF diagnosis were identified.

We also performed functional analysis on the different causes of HF across different time-

Table 3. Classification efficiency of the identified mRNA and lncRNA diagnostic biomarkers related to different etiologies in heart failure (HF)

Biomarker	Type	Combination	Data set	Accuracy	AUC
DCM-NF	mRNA	COPZ2	Train set	0.956	0.988
		CPED1	Test set-GSE1145	0.974	0.993
		TGFBR2	Test set-GSE46224	1.000	1.000
	lncRNA	AC244090.1	Train set	0.824	0.899
		SNHG3	Test set-GSE1145	0.921	0.966
		LINC01091	Test set-GSE46224	0.938	0.984
		AC004687.1*			
	LMRP	AC244090.1/COPZ2	Train set	0.732	0.779
		AC244090.1/GGT5	Test set-GSE1145	0.895	0.879
		FGD5-AS1/CPED1	Test set-GSE46224	0.750	0.859
HCM-NF	mRNA	PLAGL1	Train set	0.963	0.820
		SCG2	Test set-GSE1145	1.000	1.000
		HCLS1	Test set-GSE133054	0.875	0.922
	lncRNA	LINC01128	Train set	0.963	0.840
		LINC00894	Test set-GSE1145	0.938	1.000
		AC078850.1*	Test set-GSE133054	1.000	1.000
	LMRP	SNHG6/IFT43	Train set	0.917	0.770
		FGD5-AS1/TACC1	Test set-GSE1145	0.938	0.964
		FGD5-AS1/TGFBR2	Test set-GSE133054	0.750	0.828
		LINC00342/PLAGL1			
ICM-NF	mRNA	METRN	Train set	0.995	0.998
		EIF4EBP1	Test set-GSE1145	0.786	0.936
			Test set-GSE46224	0.813	0.891
	lncRNA	ANKRD10-IT1	Train set	0.960	0.957
		AC124798.1	Test set-GSE46224	0.750	0.859
		AC027279.1			
	LMRP	TP53TG1/UQCRB	Train set	0.955	0.979
		RAB11B-AS1/EIF4EBP1	Test set-GSE1145	0.810	0.883
			Test set-GSE46224	0.750	0.859
PPCM-NF	mRNA	IFT43	Train set	0.985	0.993
		GTF3A	Test set-GSE1145	0.867	1.000
	lncRNA	PCBP1-AS1	Train set	0.985	0.979
		LINC01128	Test set-GSE1145	0.933	1.000
		LINC01554			
	LMRP	ZFAS1/RBM47	Train set	0.959	0.988
		AL513327.1/NCBP3			
		AC002350.2/DCBLD2			

Note: '*' denotes that the lncRNA is not re-annotated in GSE1145 test set.

ordered levels and yielded several significant findings. First, we found that the functions related to energy metabolism occurred in the early stage of DCM and PPCM. In support of our finding, previous studies have demonstrated that the dysfunction to ADP/ATP carriers may cause the imbalance of intracellular energy transmission and demand, thereby promoting the occurrence of DCM [30, 31]. Hence, targeting cardiomyocyte metabolism could be a new treatment direction for non-ischemic HF [32]. For PPCM, a report indicated that the inhibition of β1AA's on PGC-1α-related pathways could impair mitochondrial energy metabolism, leading to the occurrence of PPCM [33]. Second, we found that the functional categories of actin and cytoskeletal tissue were associated with DCM and HCM. Consistently, a previous study

reported that mutations in the LMNA gene caused nuclear and cytoskeletal actin disorders by affecting Lamin A/C, which further triggered DCM initiation [34]. Similarly, the mutations in MLP were related to the development of HCM via affecting the cytoskeleton protein assembling from actin [35]. Third, we found that the 'Responses to zinc ion' were significantly enriched in ICM and PPCM, which was supported by a previous publication in which experimental data confirmed that the serum zinc level in ICM patients was significantly lower than that in normal donors [36]. Finally, we found that inflammatory response during HCM and PPCM progression might be mediated by the complement system, whereas ICM might be induced by cytokines. A comparative study verified that the increase in immune complexes concentration was associated with the decrease in complement IgG, C4, and hemolytic activity in HCM, while the complement system did not play a role in ICM [37].

Furthermore, we assessed the extent of a dysregulated LMRP with three approaches: gene expression, gene interaction, and the influence of LMRP dysregulation. Therefore, our analysis not only reflected the expression changes of a single gene and the concurrent changes between two genes, but also revealed the influence of LMRP dysregulation on the expression changes of other genes. Compared with the traditional method, our method identified more dysregulated LMRPs and acquired more HF-related biological functions. Additionally, for diagnostic biomarkers we identified, we examined the gene expression between the different etiologies with HF and normal samples in multiple test sets (Supplementary Table 10 and Supplementary Figure 1). The results demonstrated that the expression levels of some genes showed significant difference (t-test, $P < 0.05$). Since the time-series gene expression data of human cardiomyopathy, which can provide more meaningful information than the data from steady-state, is unavailable, we constructed TO-BLMRN for HF with different etiologies based on the BFS algorithm. The advantage of this method was that there was no need to correct or standardize the expression values at different time points and conditions. This method could be combined with time-series sequencing data and could detect more precise biological characteristics during disease progression.

Conclusion

In summary, the present study integrated multiple RNA-Seq datasets to construct time-ordered dysregulated lncRNA-mRNA regulation networks in different causes that induced HF and explored the common and different biological features during HF progression with different etiologies. Meanwhile, several panels of diagnostic biomarkers were identified for different causes of HF. These findings will provide new insights into the molecular mechanisms underlying the development of HF from different causes. Our findings will also provide a basis for the personalized treatment of HF patients.

Acknowledgements

This work was supported by the Natural Science Foundation of Heilongjiang Province (LH2021-F042), the Innovation Special Fund of Harbin Science and Technology Bureau of Heilongjiang Province (2017RAQXJ203), the Postdoctoral Foundation of Heilongjiang Province (LBH-Q17133). The funders had no role in study design, data collection and analysis, interpretation, and writing of the report.

Disclosure of conflict of interest

None.

Abbreviations

HF, heart failure; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy; ICM, ischemic cardiomyopathy; PPCM, peripartum cardiomyopathy; NF, non-heart failure; lncRNA, long non-coding RNA; SDE, significantly differential expressed; LMRPs, lncRNA-mRNA regulation pairs; GEO, Gene Expression Omnibus; PCC, Pearson correlation coefficients; CDF, cumulative distribution function; FC, fold change; BH, Benjamini-Hochberg; TO-BLMRN, time-ordered background lncRNA-mRNA regulation networks; BFS, breadth-first search; SVM, support vector machine; AUC, the area under the receiver operating characteristic curve; PCA, principal component analysis; DRGene, disease-related genes; GO BP, gene ontology biological process; SDs, standard deviations.

Address correspondence to: Hongbo Shi, College of Bioinformatics Science and Technology, Harbin

Medical University, No. 157 42# Baojian Road, Harbin 150076, Heilongjiang Province, China. Tel: +86-451-86669617; Fax: +86-451-86669617; E-mail: shihongbo@ems.hrbmu.edu.cn; Guangde Zhang, Department of Cardiology, The Fourth Affiliated Hospital of Harbin Medical University, No. 37 Yiyuan Road, Harbin 150010, Heilongjiang Province, China. E-mail: zhangguangde@ems.hrbmu.edu.cn

References

[1] You M, Lin M, Gong Y, Wang S, Li A, Ji L, Zhao H, Ling K, Wen T, Huang Y, Gao D, Ma Q, Wang T, Ma A, Li X and Xu F. Household fluorescent lateral flow strip platform for sensitive and quantitative prognosis of heart failure using dual-color upconversion nanoparticles. ACS Nano 2017; 11: 6261-6270.

[2] Gomes CPC, Schroen B, Kuster GM, Robinson EL, Ford K, Squire IB, Heymans S, Martelli F, Emanueli C and Devaux Y; EU-CardioRNA COST Action (CA17129). Regulatory RNAs in heart failure. Circulation 2020; 141: 313-328.

[3] Braunwald E. Heart failure. JACC Heart Fail 2013; 1: 1-20.

[4] Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr, Colvin MM, Drazner MH, Filippatos GS, Fonarow GC, Givertz MM, Hollenberg SM, Lindenfeld J, Masoudi FA, McBride PE, Peterson PN, Stevenson LW and Westlake C. 2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines and the Heart Failure Society of America. J Am Coll Cardiol 2017; 70: 776-803.

[5] Lin F, Gong X, Yu P, Yue A, Meng Q, Zheng L, Chen T, Han L, Cao H, Cao J, Liang X, Hu H, Li Y, Liu Z, Zhou X and Fan H. Distinct circulating expression profiles of long noncoding RNAs in heart failure patients with ischemic and non-ischemic dilated cardiomyopathy. Front Genet 2019; 10: 1116.

[6] Yang KC, Yamada KA, Patel AY, Topkara VK, George I, Cheema FH, Ewald GA, Mann DL and Nerbonne JM. Deep RNA sequencing reveals dynamic regulation of myocardial noncoding RNAs in failing human heart and remodeling with mechanical circulatory support. Circulation 2014; 129: 1009-1021.

[7] Li YY, Feldman AM, Sun Y and McTiernan CF. Differential expression of tissue inhibitors of metalloproteinases in the failing human heart. Circulation 1998; 98: 1728-1734.

[8] Sweet ME, Cocciolo A, Slavov D, Jones KL, Sweet JR, Graw SL, Reece TB, Ambardekar AV, Bristow MR, Mestroni L and Taylor MRG. Transcriptome analysis of human heart failure reveals dysregulated cell adhesion in dilated cardiomyopathy and activated immune pathways in ischemic heart failure. BMC Genomics 2018; 19: 812.

[9] Burke MA, Chang S, Wakimoto H, Gorham JM, Conner DA, Christodoulou DC, Parfenov MG, DePalma SR, Eminaga S, Konno T, Seidman JG and Seidman CE. Molecular profiling of dilated cardiomyopathy that progresses to heart failure. JCI Insight 2016; 1: e86898.

[10] Li M, Parker BL, Pearson E, Hunter B, Cao J, Koay YC, Guneratne O, James DE, Yang J, Lal S and O'Sullivan JF. Core functional nodes and sex-specific pathways in human ischaemic and dilated cardiomyopathy. Nat Commun 2020; 11: 2843.

[11] Sama IE, Woolley RJ, Nauta JF, Romaine SPR, Tromp J, Ter Maaten JM, van der Meer P, Lam CSP, Samani NJ, Ng LL, Metra M, Dickstein K, Anker SD, Zannad F, Lang CC, Cleland JGF, van Veldhuisen DJ, Hillege HL and Voors AA. A network analysis to identify pathophysiological pathways distinguishing ischaemic from non-ischaemic heart failure. Eur J Heart Fail 2020; 22: 821-833.

[12] Xiao Y, Xu C, Xu L, Guan J, Ping Y, Fan H, Li Y, Zhao H and Li X. Systematic identification of common functional modules related to heart failure with different etiologies. Gene 2012; 499: 332-338.

[13] Jiang Z, Guo N and Hong K. A three-tiered integrative analysis of transcriptional data reveals the shared pathways related to heart failure from different aetiologies. J Cell Mol Med 2020; 24: 9085-9096.

[14] Bolger AM, Lohse M and Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014; 30: 2114-2120.

[15] Kim D, Paggi JM, Park C, Bennett C and Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 2019; 37: 907-915.

[16] Liao Y, Smyth GK and Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014; 30: 923-930.

[17] Zhang Y, Parmigiani G and Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. NAR Genom Bioinform 2020; 2: lqaa078.

[18] Zhang X, Sun S, Pu JK, Tsang AC, Lee D, Man VO, Lui WM, Wong ST and Leung GK. Long non-coding RNA expression profiles predict clinical phenotypes in glioma. Neurobiol Dis 2012; 48: 1-8.

[19] Shi H, Sun H, Li J, Bai Z, Wu J, Li X, Lv Y and Zhang G. Systematic analysis of lncRNA and

microRNA dynamic features reveals diagnostic and prognostic biomarkers of myocardial infarction. Aging (Albany NY) 2020; 12: 945-964.
[20] Xiao Y, Hsiao TH, Suresh U, Chen HI, Wu X, Wolf SE and Chen Y. A novel significance score for gene selection and ranking. Bioinformatics 2014; 30: 801-807.
[21] Hou L, Chen M, Zhang CK, Cho J and Zhao H. Guilt by rewiring: gene prioritization through network rewiring in genome wide association studies. Hum Mol Genet 2014; 23: 2780-2790.
[22] Hotelling H. New light on the correlation coefficient and its transforms. Journal of the Royal Statistical Society 1953; 15: 193-232.
[23] Chang YM, Lin HH, Liu WY, Yu CP, Chen HJ, Wartini PP, Kao YY, Wu YH, Lin JJ, Lu MJ, Tu SL, Wu SH, Shiu SH, Ku MSB and Li WH. Comparative transcriptomics method to infer gene coexpression networks and its applications to maize and rice leaf transcriptomes. Proc Natl Acad Sci U S A 2019; 116: 3091-3099.
[24] Pinero J, Ramirez-Anguita JM, Sauch-Pitarch J, Ronzano F, Centeno E, Sanz F and Furlong LI. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res 2020; 48: D845-D855.
[25] Bao Z, Yang Z, Huang Z, Zhou Y, Cui Q and Dong D. LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. Nucleic Acids Res 2019; 47: D1034-D1037.
[26] Li J, Chen Z, Tian L, Zhou C, He MY, Gao Y, Wang S, Zhou F, Shi S, Feng X, Sun N, Liu Z, Skogerboe G, Dong J, Yao R, Zhao Y, Sun J, Zhang B, Yu Y, Shi X, Luo M, Shao K, Li N, Qiu B, Tan F, Chen R and He J. LncRNA profile study reveals a three-lncRNA signature associated with the survival of patients with oesophageal squamous cell carcinoma. Gut 2014; 63: 1700-1710.
[27] Chiu YC, Wu CT, Hsiao TH, Lai YP, Hsiao C, Chen Y and Chuang EY. Co-modulation analysis of gene regulation in breast cancer reveals complex interplay between ESR1 and ERBB2 genes. BMC Genomics 2015; 16 Suppl 7: S19.
[28] Jiang C, Li Y, Zhao Z, Lu J, Chen H, Ding N, Wang G, Xu J and Li X. Identifying and functionally characterizing tissue-specific and ubiquitously expressed human lncRNAs. Oncotarget 2016; 7: 7120-7133.
[29] Huang da W, Sherman BT and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009; 4: 44-57.
[30] Schulze K, Becker BF and Schultheiss HP. Antibodies to the ADP/ATP carrier, an autoantigen in myocarditis and dilated cardiomyopathy, penetrate into myocardial cells and disturb energy metabolism in vivo. Circ Res 1989; 64: 179-192.
[31] Schultheiss HP. Dysfunction of the ADP/ATP carrier as a causative factor for the disturbance of the myocardial energy metabolism in dilated cardiomyopathy. Basic Res Cardiol 1992; 87 Suppl 1: 311-320.
[32] Dass S, Holloway CJ, Cochlin LE, Rider OJ, Mahmod M, Robson M, Sever E, Clarke K, Watkins H, Ashrafian H, Karamitsos TD and Neubauer S. No evidence of myocardial oxygen deprivation in nonischemic heart failure. Circ Heart Fail 2015; 8: 1088-1093.
[33] Shi L, Liu J, Zhang Y, Chen M and Liu J. beta1 adrenoceptor antibodies induce myocardial apoptosis via inhibiting PGC-1alpha-related pathway. BMC Cardiovasc Disord 2020; 20: 269.
[34] Ho CY, Jaalouk DE, Vartiainen MK and Lammerding J. Lamin A/C and emerin regulate MKL1-SRF activity by modulating actin dynamics. Nature 2013; 497: 507-511.
[35] Geier C, Perrot A, Ozcelik C, Binner P, Counsell D, Hoffmann K, Pilz B, Martiniak Y, Gehmlich K, van der Ven PF, Furst DO, Vornwald A, von Hodenberg E, Nurnberg P, Scheffold T, Dietz R and Osterziel KJ. Mutations in the human muscle LIM protein gene in families with hypertrophic cardiomyopathy. Circulation 2003; 107: 1390-1395.
[36] Martin-Lagos F, Navarro-Alarcon M, Terres-Martos C, López-G de la Serrana H and Lopez-Martinez MC. Serum copper and zinc concentrations in serum from patients with cancer and cardiovascular disease. Sci Total Environ 1997; 204: 27-35.
[37] Kardaszewicz B, Rogala E, Tendera M, Kardaszewicz P and Jarzab J. Circulating immune complexes in hypertrophic cardiomyopathy and ischemic heart disease. Kardiol Pol 1991; 34: 21-24.

Supplementary Table 1. Summary of the SDE genes in HF with different etiologies

Supplementary Table 2. The identified dysregulated LMRPs in HF with different etiologies using our method

Supplementary Table 3. Summary of dysregulated LMRPs and their significantly enriched functions using the traditional method

Supplementary Table 4. Significantly enriched functions using our method

Supplementary Table 5. Summary of LMRPs at each time-ordered level in the TO-BLMRN

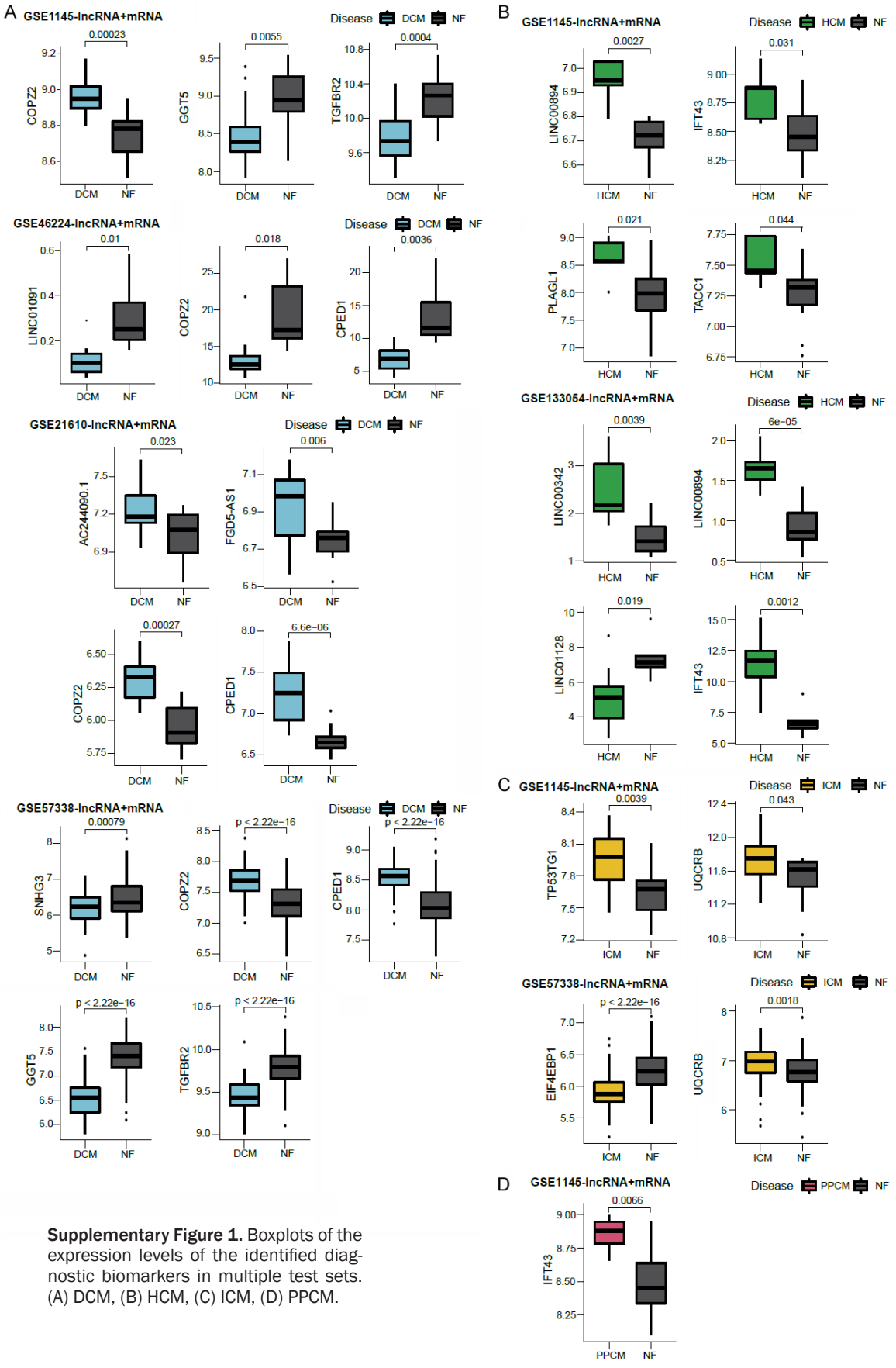
Supplementary Table 6. Significantly enriched GO BP terms of dysregulated LMRPs at each time-ordered level in DCM

Supplementary Table 7. Significantly enriched GO BP terms of dysregulated LMRPs at each time-ordered level in HCM

Supplementary Table 8. Significantly enriched GO BP terms of dysregulated LMRPs at each time-ordered level in ICM

Supplementary Table 9. Significantly enriched GO BP terms of dysregulated LMRPs at each time-ordered level in PPCM

Supplementary Table 10. Summary of the number of the identified diagnostic biomarkers in multiple test sets



Supplementary Figure 1. Boxplots of the expression levels of the identified diagnostic biomarkers in multiple test sets. (A) DCM, (B) HCM, (C) ICM, (D) PPCM.